QSAR treatment of meisoindigo derivatives as a potentbreast anticancer agent

Evie Kama Lestari[1], AgusDwi Ananto[1], Maulida Septiyana[2], Saprizal Hadisaputra[3*]

[1]Department of Pharmaceutics, University of Mataram.JalanMajapahit No 62 Mataram, 83125, Indonesia
[2]Chemistry Department, Faculty of Science, Mataram University, Indonesia
[3]Chemistry Education Divison,University of Mataram.JalanMajapahit No 62 Mataram, 83125, Indonesia
Email: rizal@unram.ac.id

ABSTRACT

A quantitative structure-activity relationship (QSAR) analysis of meisoindigo derivatives as a breast anticancer has been carried out. This study aimed to obtain the best QSAR model in order to design new meisoindigo based compounds with best anticancer activity. The semiempirical PM3 method was used for descriptor calculation. The best QSAR model was built using multilinear regression (MLR) with enter method. It was found that there were 19 new meisoindigo derivativeswith better predictive a potent anticancer agent. The best compound was *(E)-2-(1-((3-ethylisoxazol-5-yl)methyl)-2-oxoindolin-3-ylidene)-N-(4-methoxyphenyl)acetamide* with the value of IC_{50}5.31144 x10^{-15} (μM).

Keywords: QSAR, meisoindigo, anticancer.

INTRODUCTION

Cancer is an abnormal growth of cells in the body's tissues which gradually become malignant. These cells grow faster and cannot experience apoptosis. Based on the data reported in 2016 there were around 14 million cancer cases in which 8.2 million cases were reported dead [6]. The type of cancer that causes most death in women, especially those aged 40 years and over, is breast cancer.

Experimental research and development of new drugs in the laboratory requires considerable time and costs. In addition, the results obtained are likely to be unsatisfactory, so that a series of laboratory work becomes useless and wasted. Application of computational chemistry can be important in medicinal chemistry, especially for drug design, chemical properties theory, and biological activity of a molecule [8]. One of the potential compounds for anticancer is meisoindigo derivative compound. The in-vitro test for breast cancer cells shows that meisoindigo derivative compounds have good activity against these cancer cells [4].

This study will carry out a QSAR analysis of meisoindigo and its derivative compounds with the parameterized Model 3 (PM3) semiempirical method. The results of this study are the QSAR equation, which will then be used as a reference for the design of new meisoindigo derivative compounds as anticancer for abreasts cancer cells that have better biological activity.

METHODS

Molecular Modeling. Modeling of the molecular structure of the meisoindigo derivative compound was carried out using the PM3 semiempirical method and with a Convergence Limit of 0.001 kcal / Å. Optimization was carried out based on the Polak-Ribiere algorithm to achieve the minimum energy with Root Mean Square (RMS) of 0.001 kcal / Å [1]. Descriptors that will be counted represent 3 main parameters including hydrophobic, electronic, and steric parameters. The materials used were 20 types of meisoindigo derivative compounds obtained from the results of the study [4] which were known to have inhibitory activity on CDK4. The structure of the parent compound from the meisoindigo derivative can be seen in Figure 1. Furthermore, for the data of compounds with biological activities, see Table 1.
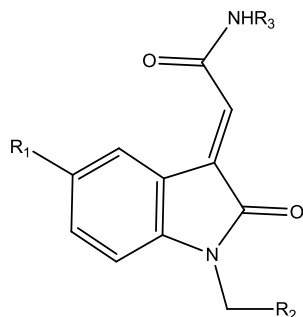
Figire 1. Meisoindigo derivative parent compound

QSAR Equation Analysis. QSAR equation analysis was done by statistical analysis using the SPSS enter method, which was conducted by entering several selected independent variables (data descriptors), that generated some equation models which were the relation of physico-chemical properties of independent variables with anticancer activities (Log $1/IC_{50}$) as a dependent variable. The models are then tested for validity [3].

Validation of QSAR Equations. The chosen QSAR model must meet all statistical criteria used and had specific descriptors that represented hydrophobic, electronic, and steric parameters. Validity testing of several equation models wasdone by calculating the value of r (correlation coefficient) and r^2 (coefficient of determination) close to 1, adjusted r^2 with the biggestvalue, smallest SE (standard error), $F_{count}/F_{table} > 1$ and PRESS (predicted recidual sums of squares) with the smallest value [10].

New Compound Design. The design of the new compound was done by modifying the type and position of the substituent. The position of the substituent was focused on the active site. The active site was chosen because it has atoms that are responsible for a breast anticancer activity from meisoindigo derivative compounds. The next step was to perform calculations using the semiempirical PM3 method according to the method used in determining the QSAR compound against all new compounds. Theoretical breast anticancer activity, Log $1/IC_{50}$ from the yield compound, was calculated using the QSAR equation that hadbeen selected. Compounds with a high Log

$1/IC_{50}$ value indicate that these compounds have high at breast anticancer activity and these compounds can be proposed for synthesis [2].

RESULTS AND DISCUSSION

Descriptors used in this study are descriptors representing electronic parameters, hydrophobic and steric. The electronic parameter consists of the net charge of the atom (q), isolated atomic energy (Eat.is), dipole (μ), hydration energy (EH), binding energy (Eb). Descriptors representing hydrophobic parameters are partition coefficient of n-octanol / water (Log P), meanwhile descriptors representing steric parameters consist of the gradient (Gr) and surface area approx (SA). The net charge of atoms used is atoms C6, C8, C9, C10, N11, C13, O14, and C15. Approx (SA) surface area descriptors, hydration energy (EH), and n-octanol / water (coefficient P) partition coefficients were obtained from the structure of the compound optimized in Hyperchem 8.0 software by selecting the next compute menu then selecting QSAR properties. Whereas for the atomic net charge (q), gradient (Gr), isolated atomic energy (Eat.is), dipole (μ), and binding energy (Eb) descriptors obtained from the saved record file in the form (file.log).

The QSAR equation can be obtained from various multivariate statistical methods that produce satisfactory results. The most widely used basic method is regression analysis. The method correlates several independent variables x (in the form of physical-chemical parameters in the Hansch method or the value of indicator variables in the Free-Wilson method) with non-independent or bound variables y (in the form of biological activity parameters of the compound). In this study, the QSAR equation was determined by regression analysis. The analysis is done by selecting the Analysis menu, then regression, and finally select linear. Determine the dependent variable, which is an experimental Log $1/IC_{50}$ value, and an independent variable, which is 16 descriptors. The next step selects the enter method and clicks ok. The results of multilinear regression analysis are presented in table 3 for 14 equation models.

Table 1. Meisoindigo derivatives used in thestudy (Chiou et al., 2015).

Compounds	R1	R2	R3	IC$_{50}$ (µM)
1	H	p-Cl-Ph	p-pyridinil	14,0
2	H	p-Cl-Ph	6-quinolinyl	3,5
3	H	3-methylisoxazolyl	p-MeO-Ph	4,0
4	H	3-methylisoxazolyl	p-pyridinil	3,6
5	H	3-methylisoxazolyl	6-quinolinyl	3,2
6	MeO	3-methylisoxazolyl	6-quinolinyl	2,3
7	H	2-naphthalenyl	6-quinolinyl	4,7
8	H	p-CF3-Ph	6-quinolinyl	1,5
9	H	p-Br-Ph	6-quinolinyl	2,5
10	H	p-NO2-Ph	6-quinolinyl	2,4
11	Me	p-Cl-Ph	6-quinolinyl	6,2
12	H	2,6-Cl2-Ph	6-quinolinyl	4,1
13	H	3-methylisoxazolyl	3-methylisoxazolyl	3,9
14	H	3-methylisoxazolyl	m-pyridinil	13,4
15	MeO	p-NH2-Ph	6-quinolinyl	27,9
16	MeO	p-Br-Ph	6-quinolinyl	3,1
17	OH	p-Br-Ph	6-quinolinyl	3,5
18	H	H	Phenyl	84,9
19	H	H	6-quinolinyl	44,7
20	H	Me	6-quinolinyl	64,3

Table 2. Training set data and test set

Compound Code	IC$_{50}$ (µM)	Compound Code	IC$_{50}$ (µM)
S8	1,5	S3	4,0
S6	2,3	S12	4,1
S10	2,4	S7	4,7
S9	2,5	S11*	6,2
S16	3,1	S14	13,4
S5*	3,2	S1*	14,0
S2	3,5	S15	27,9
S17	3,5	S19*	44,7
S4	3,6	S20	64,3
S13	3,9	S18	84,9

Compounds with a sign (*) are compounds that are used as a test set.

The selected QSAR equation model must meet all the statistical criteria used and have specific descriptors that represent each hydrophobic, electronic and steric parameter. The statistical criteria used are the value of the correlation coefficient (r), the coefficient of determination (r^2), adjusted r^2, a standard of error (SE), and F_{count}/F_{table}.The calculation results of the statistical criteria for multilinear regression analysis can be seen in table 3.

Based on data from table 3, it can be seen that the values of r and r^2 for equations 1 to 9, 13, and 14 are above 0.8, therefore the model or equation meets the requirements of a good model. The values of r and r^2 provide an illustration of the strength of the correlation between the independent variable (descriptor) and the non-independent variable (Log1/IC$_{50}$) and also determine how well the linearity level of the equation model predicts other variables.

The next analysis is the adjusted r^2 value which is the value of r^2 which has been

corrected for a number of variables. The adjusted r^2 value used is more sensitive to the estimation of errors found in each equation. From table 3 it can be seen that the value of model 14 has the largest adjusted r^2 value which indicates that model 14 has a good accuracy in describing the relationship between independent and non-independent variables. This is also supported by the standard of error data where model 14 has a relatively smaller error compared to other models.

The other statistical criteria are F_{count}/F_{tabel}. The F_{count}/F_{tabel} value is able to describe the significance level of the data with a confidence level of 95%. The greater the F_{count}/F_{tabel} value, the more significant the data will be. Since the 6-12 model has a value of $F_{count}/F_{tabel}> 1$, it is assumed that all of these models are statistically acceptable. While models 4, 5, 13 and 14 have F_{count}/F_{tabel} values close to 1 which indicates that the models are close to significant. Based on the analysis of statistical criteria from r; r^2; adjusted r^2; SE; F_{count}/F_{tabel}; and the PRESS test, the models 4, 13 and 14 are considered to be the best QSAR anticancer equation model of meisoindigo derivative compounds.

The selection of the best QSAR models from the three models was carried out by validation tests using four predetermined test sets. Validation test was done by predicting a breast anticancer activity using three selected equation models, then making a log$1/IC_{50}$ relationship curve experiment with prediction $1/IC_{50}$ as seen in table 4 and in figure 2.

The relationship between the comparison of Log $1/IC_{50}$ experiments with predictive Log $1/IC_{50}$ will produce slope values and r^2 values that can be taken into consideration in determining the best equation model. If the predicted value of Log $1/IC_{50}$ approaches $1/IC_{50}$ experiment, the slope value on the curve will get closer to one. Therefore to determine the QSAR equation model is best done by finding the slope value that is closest to one. Based on the comparison curve between predictions and experiments, it can be seen that the slope value that is closest to one is model 14, with a value of 0.821. While for models 4

and 13 give a slope value of 0.411 and 0.632, respectively.

The value of r^2 which is close to 1 indicates that the correlation between the independent variable and the dependent variable is significant but if the value of r^2 approaches 0 then the correlation between these variables is not significant. Based on the curve, it can be seen that the value of r^2 which is closest to one is model 14, which is 0.994. Whereas, models 4 and 13 have r^2 values of 0.399 and 0.722. Thus, model 14 is chosen as the best equation.

New Compound Design. Model 14 as the chosen model was used as a guiding model in designing and predicting new breast anticancer activities from meisoindigo derivative compounds. In designing new compounds, the selection of the structure of the main compound is based on the meisoindigo derivative compound which has the best breast anticancer activity. The design of the compound begins by modifying the type and position of the substituent. There are 65 new compounds that have been designed and then modeled.

Descriptors of new compounds obtained from these calculations are then included in the best QSAR equation model, namely model 14, to calculate breast anticancer activity theoretically. The calculation results obtained 19 design compounds that have better activity or lower IC_{50} value compared to that of experiments. It can be seen in table 4.4.

The higher the Log $1/IC_{50}$ value of a compound, or the smaller the IC_{50} value, the better the inhibitory ability of the compound. Among the 19 compounds, the design compound which has the smallest IC_{50} value is found in compound number 25.

Table 4. The QSAR equation model results from multilinear regression analysis with the enter method.

Model	Variable	R	r^2	Adjusted r^2	SE	F_{hitung}	F_{tabel}	F_{hit}/F_{tab}	PRESS
1	qC13, Gr, qN11, Eat.is, µ, qC8, Log P, EH, Eb, qC9, SA, qO10	0,925	0,856	0,280	0,446	1,485	8,745	0,170	2188,854478
2	qC13, qN11, Eat.is, µ, qC8, Log P, EH, Eb, qC9, SA, qO10	0,923	0,852	0,446	0,391	2,098	5,936	0,353	655,0808156
3	qC13, qN11, Eat.is, µ, qC8, Log P, EH, Eb, qC9, qO10	0,920	0,847	0,541	0,356	2,768	4,735	0,585	486,5321111
4	qC13, qN11, Eat.is, µ, qC8, Log P, EH, qC9, qO10	0,913	0,834	0,586	0,338	3,360	4,099	0,820	0,002685139
5	qC13, qN11, Eat.is, µ, qC8, Log P, qC9, qO10	0,900	0,810	0,592	0,335	3,723	3,726	0,999	$2,28906 \times 10^{-5}$
6	qC13, qN11, Eat.is, µ, Log P, qC9, qO10	0,889	0,790	0,607	0,329	4,309	3,500	1,231	$1,21488 \times 10^{-5}$
7	qC13, qN11, Eat.is, µ, qC9, qO10	0,866	0,751	0,585	0,338	4,518	3,374	1,339	$3,86559 \times 10^{-5}$
8	qC13, Eat.is, µ, qC9, qO10	0,815	0,664	0,496	0,373	3,956	3,326	1,189	$1,09203 \times 10^{-5}$
9	Eat.is, µ, qC9, qO10	0,804	0,646	0,517	0,365	5,010	3,357	1,493	$2,0727 \times 10^{-7}$
10	Eat.is, qC9, qO10	0,798	0,636	0,546	0,354	7,001	3,490	2,006	$6,32813 \times 10^{-5}$
11	Eat.is, qO10	0,789	0,623	0,565	0,346	10,731	3,806	2,820	$3,76452 \times 10^{-5}$
12	Eat.is	0,749	0,562	0,530	0,360	17,940	4,600	3,900	$2,6014 \times 10^{-6}$
13	EH, µ, qN11, Eat.is, qO10, qC9, qC15, Log P, qC13	0,927	0,859	0,647	0,312	4,055	4,099	0,989	0,000471666
14	qC6, qO14, qN11, µ, Eat.is, Log P, qC15, EH, qC9, qC13, qO10	0,968	0,937	0,764	0,255	5,421	5,936	0,913	0,004559995

Evie Kama Lestari et al

Table 3. Data comparison of experimental Log 1 / IC_{50} values with predictive Log 1 / IC_{50} values

Test Compound	Log 1/IC_{50} experiment	Log 1/IC_{50} prediction		
		Model 4	Model 13	Model 14
1	4,8539	4,4924	4,6232	4,8854
5	5,4949	5,0564	5,1131	5,4509
11	5,2076	5,2725	5,2876	5,1475
19	4,3497	4,7319	4,5368	4,4934
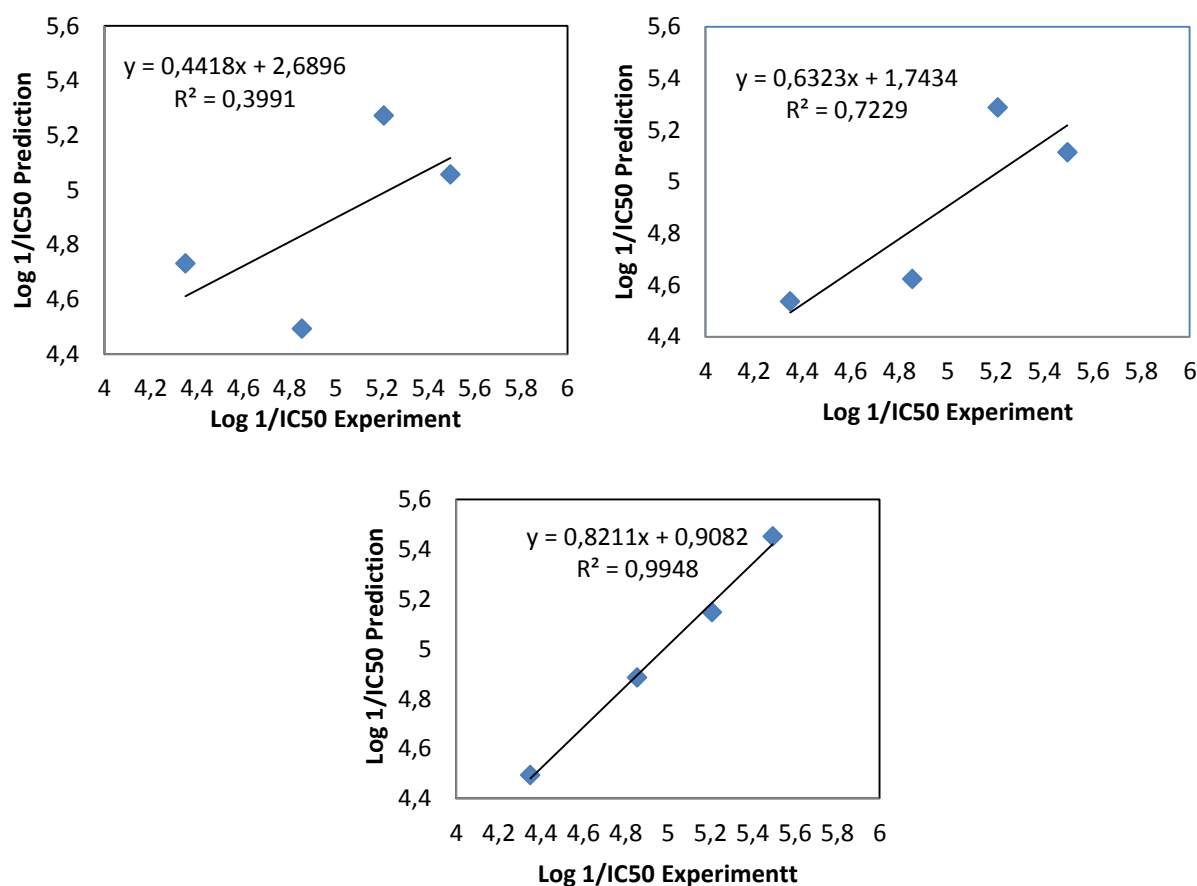
Figure 2.Log 1 / IC50 relationship curve experiments with Log 1 / IC50 predictions for models 4, 13 and 14.

Evie Kama Lestari et al

Table 4. Design compounds that have better activity than experiments

No	Compound	IUPAC name	IC$_{50}$ (µM)
1	23	*(E)-N-(3-methoxyphenyl)-2-(1-((3-methylisoxazol-5-yl)methyl)-2-oxoindolin-3-ylidene)acetamide*	5.59392x10^{-8}
2	25	*(E)-2-(1-((3-ethylisoxazol-5-yl)methyl)-2-oxoindolin-3-ylidene)-N-(4-methoxyphenyl)acetamide*	5.31144 x10^{-15}
3	30	*(E)-2-(1-((3-methylisoxazol-5-yl)methyl)-2-oxoindolin-3-ylidene)-N-(quinolin-5-yl)acetamide*	4.01254 x10^{-8}
4	33	*(E)-2-(1-((3-ethylisoxazol-5-yl)methyl)-2-oxoindolin-3-ylidene)-N-(quinolin-5-yl)acetamide*	3.25918 x10^{-8}
5	38	*(E)-2-(2-oxo-1-(3-(trifluoromethyl)benzyl)indolin-3-ylidene)-N-(quinolin-6-yl)acetamide*	0.85691263
6	39	*(E)-2-(2-oxo-1-(2-(trifluoromethyl)benzyl)indolin-3-ylidene)-N-(quinolin-6-yl)acetamide*	0.350508368
7	40	*(E)-2-(2-oxo-1-(4-(trifluoromethyl)benzyl)indolin-3-ylidene)-N-(quinolin-5-yl)acetamide*	8.57865 x10^{-7}
8	41	*(E)-2-(2-oxo-1-(4-(trifluoromethyl)benzyl)indolin-3-ylidene)-N-(quinolin-7-yl)acetamide*	1.374101302
9	42	*(E)-2-(2-oxo-1-(3-(trifluoromethyl)benzyl)indolin-3-ylidene)-N-(quinolin-7-yl)acetamide*	0.546881356
10	43	*(E)-2-(2-oxo-1-(2-(trifluoromethyl)benzyl)indolin-3-ylidene)-N-(quinolin-7-yl)acetamide*	0.101931356
11	44	*(E)-2-(2-oxo-1-(3-(trifluoromethyl)benzyl)indolin-3-ylidene)-N-(quinolin-5-yl)acetamide*	5.02664 x10^{-7}
12	45	*(E)-2-(2-oxo-1-(2-(trifluoromethyl)benzyl)indolin-3-ylidene)-N-(quinolin-5-yl)acetamide*	2.20424 x10^{-6}
13	68	*(E)-2-(1-((3-hydroxyisoxazol-5-yl)methyl)-2-oxoindolin-3-ylidene)-N-(4-methoxyphenyl)acetamide*	0.002424643
14	74	*(E)-2-(2-oxo-1-(quinolin-6-ylmethyl)indolin-3-ylidene)-N-(4-(trifluoromethyl)phenyl)acetamide*	1.415227631
15	77	*(E)-2-(5-fluoro-1-(4-nitrobenzyl)-2-oxoindolin-3-ylidene)-N-(quinolin-6-yl)acetamide*	1.181599818
16	79	*(E)-N-(3-fluoroisoxazol-5-yl)-2-(1-((3-methylisoxazol-5-yl)methyl)-2-oxoindolin-3-ylidene)acetamide*	0.279148131
17	80	*(E)-N-(3-hydroxyisoxazol-5-yl)-2-(1-((3-methylisoxazol-5-yl)methyl)-2-oxoindolin-3-ylidene)acetamide*	0.153112273
18	81	*(E)-N-(3-aminoisoxazol-5-yl)-2-(1-((3-methylisoxazol-5-yl)methyl)-2-oxoindolin-3-ylidene)acetamide*	0.079535658
19	84	*(E)-2-(1-(4-bromobenzyl)-5-cyanamido-2-oxoindolin-3-ylidene)-N-(quinolin-6-yl)acetamide*	0.132459366

There are several factors for selecting the proposed compound for analysis besides paying attention to the IC$_{50}$ value which is of small value. Compound number 25 forms a structure similar to compound 8 which experimentally has the best activity. Based on its substituents, the methoxy group (p-MeO-Ph) in R3 is a group that is an electron withdrawing that is stronger than hydrogen by showing a negative induction effect, so it is used to increase electronic parameters that play a role in drug-receptor interactions that can later provide a biological response. The substitution of the methoxy group is carried out in the *-para* position in order to influence the steric properties of the drug molecules so that it can prolong the action of the drug in the body [11].

CONCLUSIONS

The best QSAR equation model of meisoindigo derivative as anticancer in breast cancer using semiempirical PM3 method is as follows:

Log $1/IC_{50}$ = 44,316-(2.820E-5 Eis.at)-(1,734 qC9)+(25.711 qO10)+(94,825 qC13)+(7,309 qN11)-(0,257 μ)-(0.054 Log P)+(5.866 qC15)+(0,014 EH)+(58,794 qO14)-(7,241 qC6) r = 0.968 ; r^2=0.937 ; *Adjusted* r^2=0.764 ; SE=0.255 ; F_{hit}/F_{tab}= 0.913 ; PRESS=0.004559995.

There are 19 compounds of meisoindigo derivative obtained from the best QSAR equation model. The 19 compunds have better anticancer activity predictions than those of experiments, and the proposed compound is (*E*)-2-(1-((3-ethylisoxazol-5-yl)methyl)-2-oxoindolin-3-ylidene) -N- (4-methoxyphenyl) acetamide with IC_{50} value of 5.31144 x10-15.

Refrences

[1] Podunavac-Kuzmanović, S. O., & Cvetković, D. D. (2011). QSAR modeling of antibacterial activity of some benzimidazole derivatives. *Chemical Industry and Chemical Engineering Quarterly*, *17*(1), 33-38.

[2] Naik, P. K., Sindhura, Singh, T., & Singh, H. (2009). Quantitative structure–activity relationship (QSAR) for insecticides: development of predictive in vivo insecticide activity models. *SAR and QSAR in Environmental Research*, *20*(5-6), 551-566.

[3] Ye, B., Xiong, X., Deng, X., Gu, L., Wang, Q., Zeng, Z., ... & Wang, Y. (2017). Meisoindigo, but not its core chemical structure indirubin, inhibits zebrafish interstitial leukocyte chemotactic migration. *Pharmaceutical biology*, *55*(1), 673-679.

[4] Chiou, C. T., Lee, W. C., Liao, J. H., Cheng, J. J., Lin, L. C., Chen, C. Y., ... & Li, W. T. (2015). Synthesis and evaluation of 3-ylideneoxindole acetamides as potent anticancer agents. *European journal of medicinal chemistry*, *98*, 1-12.

[5] Kemenkes RI. 2015. Infodatin Cancer. Indonesian Ministry of Health Data and Information Center Cancer Situation. Jakarta :KementerianKesehatan RI.

[6] Kemenkes RI. 2016. Infodatin Center for Data and Information of the Indonesian Ministry of Health Breast Cancer.Jakarta :KementerianKesehatan RI.

[7] Oruç, E. E., Rollas, S., Kandemirli, F., Shvets, N., & Dimoglo, A. S. (2004). 1, 3, 4-thiadiazole derivatives. Synthesis, structure elucidation, and structure− antituberculosis activity relationship investigation. *Journal of medicinal chemistry*, *47*(27), 6760-6767.

[8] Hirjani, H., Mudasir, M., & Pranowo, H. D. (2018). Prediction of High Performance Liquid Chromatography Retention Time for Some Organic Compounds Based on Ab initio QSPR Study. *Acta Chimica Asiana*, *1*(1), 24-29.

[9] Mingxin, Z., Yan, L., Hongbo, W., Jianhua, Z., Hongyan, L., He, L., ... & Xiaoguang, C. (2008). The antitumor activity of meisoindigo against human colorectal cancer HT-29 cells in vitro and in vivo. *Journal of Chemotherapy*, *20*(6), 728-733.

[10] Toropov, A. A., Toropova, A. P., Puzyn, T., Benfenati, E., Gini, G., Leszczynska, D., & Leszczynski, J. (2013). QSAR as a random event: modeling of nanoparticles uptake in PaCa2 cancer cells. *Chemosphere*, *92*(1), 31-37.

[11] Gao, H., Katzenellenbogen, J. A., Garg, R., & Hansch, C. (1999). Comparative QSAR analysis of estrogen receptor ligands. *Chemical Reviews*, *99*(3), 723-744.